(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 107211000 B
(45)授权公告日 2020.06.05

(21)申请号 201580064576.7

(22)申请日 2015.11.24

(65)同一申请的已公布的文献号
申请公布号 CN 107211000 A

(43)申请公布日 2017.09.26

(30)优先权数据
62/084,656 2014.11.26 US

(85)PCT国际申请进入国家阶段日
2017.05.26

(86)PCT国际申请的申请数据
PCT/US2015/062260 2015.11.24

(87)PCT国际申请的公布数据
WO2016/085897 EN 2016.06.02

(73)专利权人 里德爱思唯尔股份有限公司雷克萨斯尼克萨斯分公司
地址 美国俄亥俄州

(72)发明人 W·基尔加隆

(74)专利代理机构 上海专利商标事务所有限公司 31100
代理人 李玲

(51)Int.Cl.
H04L 29/06(2006.01)

(56)对比文件
US 2005251865 A1,2005.11.10,
US 2003097594 A1,2003.05.22,
US 2007033079 A1,2007.02.08,
US 2006123462 A1,2006.06.08,
CN 103403685 A,2013.11.20,
CN 101809570 A,2010.08.18,
CN 104090924 A,2014.10.08,
US 2014047551 A1,2014.02.13,

审查员 陈丽锋

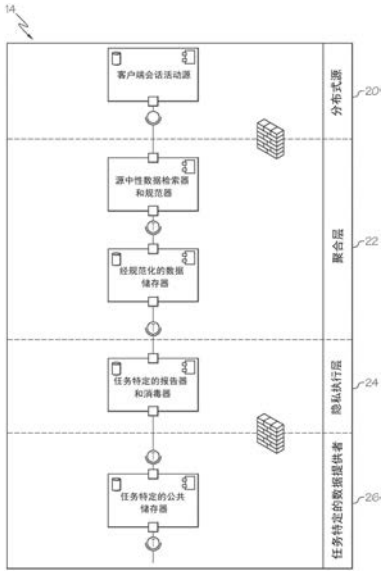
权利要求书5页 说明书10页 附图8页

(54)发明名称

用于实现隐私防火墙的系统和方法

(57)摘要

披露了用于保护隐私防火墙后面的私有数据的系统和方法。一种用于实现隐私防火墙从而确定并提供来自私有电子数据的非私有信息的系统包括数据储存库、处理设备、以及非瞬态处理器可读存储介质。所述存储介质包括编程指令，所述编程指令当被执行时使得所述处理设备：分析私有电子数据语料库从而识别所述数据的具有非私有信息的第一一个或多个部分以及所述数据的具有私有信息的第二一个或多个部分，将所述数据的所述第一一个或多个部分标记为被允许使用，判定所述数据的所述第二一个或多个部分是否包括非私有元素，并且如果所述数据的所述第二一个或多个部分包括非私有元素，则提取所述非私有元素并将所述非私有元素标记为被允许使用的信息。



CN 107211000 B

1.一种用于实现隐私防火墙以确定并从私有电子数据向位于所述隐私防火墙外的远程计算机提供非私有信息的系统,所述系统包括:

数据储存库,所述数据储存库包括私有电子数据语料库;

处理设备;以及

非瞬态处理器可读存储介质,其中,所述非瞬态处理器可读存储介质包括一条或多条编程指令,当所述一条或多条编程指令被执行时使得所述处理设备:

接收来自位于所述隐私防火墙外的所述远程计算机的查询,所述查询包括对访问所述私有电子数据语料库的一个或多个部分的请求;

分析所述隐私防火墙内的所述私有电子数据语料库,从而识别包括非私有信息的所述数据的第一一个或多个部分以及包括私有信息的所述数据的第二一个或多个部分;

将所述数据的所述第一一个或多个部分标记为被允许在所述隐私防火墙外部使用;

确定所述数据的所述第二一个或多个部分包括非私有元素,其中所述非私有元素是所述私有信息或从所述私有信息推出的附加数据的非私有的部分,且其中所述数据的所述第二一个或多个部分中的至少一个包括具有低于粒度阈值的时间戳的一个或多个元素;

从所述数据的所述第二一个或多个部分提取所述非私有元素,其中其余所述私有元素是未提取的元素;

将所述非私有元素标记为被允许在所述隐私防火墙外部使用的信息;

将来自所述数据的所述第二一个或多个部分的所述未提取的元素标记为被拒绝在所述隐私防火墙外部使用;

将具有低于所述粒度阈值的时间戳的所述一个或多个元素调整至高于所述粒度阈值的时间范围;

将一个或多个经调整元素标记为被允许在所述隐私防火墙外部使用的非私有元素;以及

向位于所述隐私防火墙外的所述远程计算机提供包括非私有信息的所述数据的所述第一一个或多个部分和所述非私有元素中的一个或多个。

2.如权利要求1所述的系统,其中,当一条或多条编程指令被执行时,使得所述处理设备分析所述私有电子数据语料库,进一步使得所述处理设备:

判定所述私有电子数据语料库是否是从已排除列表上的源获得;并且

如果所述私有电子数据语料库是从所述已排除列表上的源获得,则将所述私有电子数据语料库标记为被拒绝在所述隐私防火墙外部使用。

3.如权利要求1所述的系统,其中,当一条或多条编程指令被执行时,使得所述处理设备判定所述数据的所述第二一个或多个部分,进一步使得所述处理设备:

判定所述数据的所述第二一个或多个部分是否包括已经在其他位置出现的一个或多个元素;并且

如果所述一个或多个子部分已经出现在其他位置,则将所述元素标记为非私有元素。

4.如权利要求1所述的系统,其中,当一条或多条编程指令被执行时,使得所述处理设备判定所述数据的所述第二一个或多个部分是否包括非私有元素,进一步使得所述处理设备:

判定所述数据的所述第二一个或多个部分是否包括已经从阈值数量个不同源出现的

一个或多个元素;并且

如果所述一个或多个元素已经从至少所述阈值数量个不同源出现,则将所述一个或多个元素标记为非私有元素。

5.如权利要求1所述的系统,其中,当一条或多条编程指令被执行时,使得所述处理设备判定所述数据的所述第二一个或多个部分是否包括非私有元素,进一步使得所述处理设备:

判定所述数据的所述第二一个或多个部分是否包括包含了已经是公共知识的信息的一个或多个元素;并且

如果所述一个或多个元素包含已经是公共知识的信息,则将所述一个或多个元素标记为非私有元素。

6.如权利要求1所述的系统,其中,当一条或多条编程指令被执行时,使得所述处理设备判定所述数据的所述第二一个或多个部分,进一步使得所述处理设备:

判定所述数据的所述第二一个或多个部分是否包括较复杂的一个或多个元素,使得所述一个或多个元素可被要求为专用信息;并且

如果所述一个或多个元素较复杂,则将所述一个或多个元素标记为被拒绝在所述隐私防火墙外部使用。

7.如权利要求1所述的系统,其中,当一条或多条编程指令被执行时,使得所述处理设备判定所述数据的所述第二一个或多个部分是否包括非私有元素,进一步使得所述处理设备:

判定所述数据的所述第二一个或多个部分是否包括具有可识别序列的一个或多个子部分;并且

如果所述一个或多个子部分具有可识别序列,则将所述一个或多个子部分标记为被拒绝分布在所述隐私防火墙外部。

8.如权利要求1所述的系统,其中,当一条或多条编程指令被执行时,使得所述处理设备判定所述数据的所述第二一个或多个部分是否包括非私有元素,进一步使得所述处理设备:

判定所述数据的所述第二一个或多个部分是否包括具有精确时间戳的一个或多个元素;并且

如果所述一个或多个元素具有精确时间戳,则将所述一个或多个元素标记为被拒绝在所述隐私防火墙外部使用。

9.如权利要求1所述的系统,其中,当一条或多条编程指令被执行时,使得所述处理设备判定所述数据的所述第二一个或多个部分是否包括非私有元素,进一步使得所述处理设备:

判定所述数据的所述第二一个或多个部分是否包括具有低于粒度阈值的地理位置的一个或多个元素;并且

如果所述地理位置低于所述粒度阈值,则:

将所述一个或多个元素调整至高于所述粒度阈值的地理位置,并且

将所述一个或多个经调整元素标记为非私有元素。

10.如权利要求1所述的系统,其中,当被执行时使得所述处理设备向所述远程计算机

提供所述数据的所述第一一个或多个部分和所述非私有元素中的所述一个或多个的所述一条或多条编程指令进一步使得所述处理设备：

通过用户界面向用户提供对所述查询的响应，其中，对所述查询的所述响应包含从包括非私有信息的所述数据的所述第一一个或多个部分和所述非私有元素中的一项或多项所获得的信息。

11.一种用于实现隐私防火墙以确定并从私有电子数据向位于所述隐私防火墙外的远程计算机提供非私有信息的方法，所述方法包括：

接收来自位于所述隐私防火墙外的所述远程计算机的查询，所述查询包括对访问所述私有电子数据语料库的一个或多个部分的请求；

由处理设备分析所述隐私防火墙内的包含在储存库中的私有电子数据语料库来识别包括非私有信息的所述数据的第一一个或多个部分以及包括私有信息的所述数据的第二一个或多个部分；

由所述处理设备将所述数据的所述第一一个或多个部分标记为被允许在所述隐私防火墙外部使用；

由所述处理设备判定所述数据的所述第二一个或多个部分是否包括非私有元素，其中所述非私有元素是所述私有信息或从所述私有信息推出的附加数据的非私有的部分，且其中所述数据的所述第二一个或多个部分中的至少一个包括具有低于粒度阈值的时间戳的一个或多个元素；并且

由所述处理设备从所述数据的所述第二一个或多个部分提取所述非私有元素，其中其余所述私有元素是未提取的元素；

由所述处理设备将所述非私有元素标记为被允许在所述隐私防火墙外部使用的信息；

由所述处理设备将来自所述数据的所述第二一个或多个部分的所述未提取的元素标记为被拒绝在所述隐私防火墙外部使用；

由所述处理设备将具有低于所述粒度阈值的时间戳的所述一个或多个元素调整至高于所述粒度阈值的时间范围；

由所述处理设备将一个或多个经调整元素标记为被允许在所述隐私防火墙外部使用的非私有元素；以及

由所述处理设备向位于所述隐私防火墙外的所述远程计算机提供包括非私有信息的所述数据的所述第一一个或多个部分和所述非私有元素中的一个或多个。

12.如权利要求11所述的方法，其中，分析所述私有电子数据语料库包括：

由所述处理设备判定所述私有电子数据语料库是否是从已排除列表上的源获得；并且

如果所述数据语料库是从所述已排除列表上的源获得，则由所述处理设备将所述私有电子数据语料库标记为被拒绝在所述隐私防火墙外部使用。

13.如权利要求11所述的方法，其中，判定所述数据的所述第二一个或多个部分是否包括非私有元素包括：

由所述处理设备判定所述数据的所述第二一个或多个部分是否包括具有低于粒度阈值的地理位置的一个或多个元素；并且

如果所述地理位置低于所述粒度阈值，则：

由所述处理设备将所述一个或多个元素调整至高于所述粒度阈值的地理位置，并且

由所述处理设备将所述一个或多个经调整元素标记为非私有元素。

14.如权利要求11所述的方法，其中，向所述远程计算机提供所述数据的所述第一一个或多个部分和所述非私有元素中的所述一个或多个包括：

由所述处理设备通过用户界面向用户提供对所述查询的响应，其中，对所述查询的所述响应包含从包括非私有信息的所述数据的所述第一一个或多个部分和所述非私有元素中的一项或多项所获得的信息。

15.一种用于实现隐私防火墙以确定并从私有电子数据向位于所述隐私防火墙外的远程计算机提供非私有信息的方法，所述方法包括：

由处理设备分析包含在储存库中的私有电子数据语料库来识别包括非私有信息的所述数据的第一一个或多个部分以及包括私有信息的所述数据的第二一个或多个部分；

由所述处理设备将所述数据的所述第一一个或多个部分标记为被允许在所述隐私防火墙外部使用；

由所述处理设备判定所述数据的所述第二一个或多个部分是否包括非私有元素；

如果所述数据的所述第二一个或多个部分包括非私有元素，则：

由所述处理设备提取所述非私有元素，并且

由所述处理设备将所述非私有元素标记为被允许在所述隐私防火墙外部使用的信息；以及

由所述处理设备向所述远程计算机提供包括非私有信息的所述数据的所述第一一个或多个部分和所述非私有元素中的一个或多个，

其中，判定所述数据的所述第二一个或多个部分是否包括非私有元素包括：

由所述处理设备判定所述数据的所述第二一个或多个部分是否包括已经在其他位置出现的一个或多个元素；

如果所述一个或多个子部分已经出现在其他位置，则由所述处理设备将所述元素标记为非私有元素；

由所述处理设备判定所述数据的所述第二一个或多个部分是否包括已经从阈值数量个不同源出现的一个或多个元素；

如果所述一个或多个元素已经从至少所述阈值数量个不同源出现，则由所述处理设备将所述一个或多个元素标记为非私有元素；

由所述处理设备判定所述数据的所述第二一个或多个部分是否包括包含了已经是公共知识的信息的一个或多个元素；

如果所述一个或多个元素包含已经是公共知识的信息，则由所述处理设备将所述一个或多个子部分标记为非私有元素；

由所述处理设备判定所述数据的所述第二一个或多个部分是否包括一个或多个元素，所述一个或多个元素较复杂以使得所述一个或多个元素可要求为专用信息；

如果所述一个或多个元素较复杂，则由所述处理设备将所述一个或多个元素标记为被拒绝在所述隐私防火墙外部使用；

由所述处理设备判定所述数据的所述第二一个或多个部分是否包括具有可识别序列的一个或多个元素；

如果所述一个或多个元素具有可识别序列，则由所述处理设备将所述一个或多个元素

标记为被拒绝在所述隐私防火墙外部使用；

由所述处理设备判定所述数据的所述第二一个或多个部分是否包括具有精确时间戳的一个或多个元素；并且

如果所述一个或多个元素具有精确时间戳，则由所述处理设备将所述一个或多个元素标记为被拒绝在所述隐私防火墙外部使用。

16.一种用于通过隐私防火墙从私有电子数据向位于所述隐私防火墙外的远程计算机提供非私有信息的系统，所述系统包括：

所述隐私防火墙后面的数据储存库，所述数据储存库包括：已经被标记为非私有信息的私有电子数据语料库、不被分发的私有信息、私有信息的会被分发的非私有元素、或已去私有化的私有信息，其中所述非私有元素是所述私有信息或从所述私有信息推出的附加数据的非私有的部分；

处理设备；以及

非瞬态处理器可读存储介质，其中，所述非瞬态处理器可读存储介质包括一条或多条编程指令，当所述一条或多条编程指令被执行时，使得所述处理设备：

从位于所述隐私防火墙外部的所述远程计算机接收搜索字符串，其中，所述搜索字符串包括问题；

在所述数据储存库中搜索所述私有电子数据的与所述搜索字符串相对应的一个或多个部分；

如果所述电子数据的所述一个或多个部分包含：所述非私有信息、所述私有信息的会被分发的非私有元素、或所述已去私有化的私有信息，则提供对所述远程计算机的响应，其中，所述响应包含所述电子数据的一个或多个部分内所包含的信息，所述信息包含：所述非私有信息、所述私有信息的会被分发的非私有元素、或所述已去私有化的私有信息；

将不是所述非私有信息、所述私有信息的会被分发的非私有元素、或所述已去私有化的私有信息的其余所述私有信息标记为被拒绝使用，

其中，所述私有信息的所述非私有元素包括具有低于粒度阈值的时间戳的一个或多个元素，以及

其中，提供所述响应包括将所述一个或多个元素调整至具有高于所述粒度阈值的时间范围并提供一个或多个经调整元素。

用于实现隐私防火墙的系统和方法

[0001] 相关申请的交叉引用

[0002] 本申请要求2014年11月26日提交的题为“SYSTEMS AND METHODS FOR DATA PRIVACY FIREWALL(用于数据隐私防火墙的系统和方法)”的美国临时专利申请序列号62/084,656的权益,所述美国临时专利申请的全部公开内容通过引用结合在此。

技术领域

[0003] 本说明书总体上涉及提供隐私防火墙以保护私有和敏感数据,并且更具体地,涉及用于提供对从私有和敏感数据获得的非私有信息的访问的系统和方法。

背景技术

[0004] 目前,可以通过将与数据有关的用户活动匿名化并暴露所述已匿名化数据以便浏览和使用来提供对私有和/或敏感数据的访问。然而,此类方法可能无法有效地保护隐私,因为所述数据可以被操纵以从中提取私有信息。此类方法还可能由于数据被匿名化到不再有用的程度而失败。

[0005] 相应地,存在对以下系统和方法的需要,所述系统和方法不对私有数据匿名化,而是从中发现并提取非私有元素,其方式为使得不破坏所述数据的隐私,但私有数据中所包含的数据是有用的。

发明内容

[0006] 在一个实施例中,一种用于实现隐私防火墙从而确定并提供来自私有电子数据的非私有信息的系统包括具有私有电子数据语料库的数据储存库、处理设备、以及非瞬态处理器可读存储介质。所述非瞬态计算机可读存储介质包括一条或多条编程指令,所述一条或多条编程指令当被执行时使得所述处理设备:分析所述电子数据语料库从而识别所述数据的具有非私有信息的第一一个或多个部分以及所述数据的具有私有信息的第二一个或多个部分,将所述数据的所述第一一个或多个部分标记为被允许在所述隐私防火墙外部使用,判定所述数据的所述第二一个或多个部分是否包括非私有元素,并且如果所述数据的所述第二一个或多个部分包括非私有元素,则提取所述非私有元素并将所述非私有元素标记为被允许在所述隐私防火墙外部使用的信息。

[0007] 在另一实施例中,一种用于实现隐私防火墙从而确定并提供来自私有电子数据的非私有信息的方法包括:由处理设备分析包含在储存库中的私有电子数据语料库从而识别所述数据的包括非私有信息的第一一个或多个部分以及所述数据的包括私有信息的第二一个或多个部分;由所述处理设备将所述数据的所述第一一个或多个部分标记为被允许在所述隐私防火墙外部使用;由所述处理设备判定所述数据的所述第二一个或多个部分是否包括非私有元素;以及如果所述数据的所述第二一个或多个部分包括非私有元素,则由所述处理设备提取所述非私有元素并且由所述处理设备将所述非私有元素标记为被允许在所述隐私防火墙外部使用的信息。

[0008] 在又另一实施例中，一种用于通过隐私防火墙提供来自私有电子数据的非私有信息的系统包括：所述隐私防火墙后面的数据储存库、处理设备、以及非瞬态处理器可读存储介质。所述数据储存库包括：已经被标记为非私有信息的私有电子数据语料库、将不分发的私有信息、有待分发的私有信息的非私有元素、或已去私有化的私有信息。所述非瞬态处理器可读存储介质包括一条或多条编程指令，所述一条或多条编程指令当被执行时使得所述处理设备：从所述隐私防火墙外部的用户接收搜索字符串；在所述数据储存库中搜索所述电子数据的与所述搜索字符串相对应的一个或多个部分；并且如果所述电子数据的所述一个或多个部分包含所述非私有信息、所述有待分发的私有信息的所述非私有元素、或所述已去私有化的私有信息，则提供对所述搜索字符串的响应。所述搜索字符串包括问题，并且所述响应包含：包含所述非私有信息、所述有待分发的私有信息的所述非私有元素、或所述已去私有化的私有信息的所述电子数据的所述一个或多个部分中所包含的信息。

[0009] 鉴于以下具体描述结合附图将更完整地理解由在此描述的实施例提供的这些和附加特征。

附图说明

[0010] 附图中阐明的实施例在本质上是说明性的且示例性的并且并不旨在限制由权利要求书限定的主题。当结合以下附图阅读时，能够理解说明性实施例的以下详细描述，其中，相同的结构用相同的参考标号指示，并且在附图中：

[0011] 图1根据本文所示出和所描述的一个或多个实施例描绘了针对用于提供对隐私防火墙后面数据的访问的系统的说明性计算网络的示意性描绘；

[0012] 图2描绘了来自图1的服务器计算设备的示意性描绘，进一步展示了根据本文所示出和所描述的一个或多个实施例可以用于提供数据的硬件和软件；

[0013] 图3根据在此示出和描述的一个或多个实施例描绘了来自图1的隐私防火墙的各个层的示意性描绘；

[0014] 图4根据在此示出和描述的一个或多个实施例描绘了描绘了响应于请求而提供数据的示意性方法的流程图；

[0015] 图5根据在此示出和描述的一个或多个实施例描绘了图形用户界面的示意性搜索输入屏幕的示意性描绘；

[0016] 图6根据在此示出和描述的一个或多个实施例描绘了包含自动完成选项的图形用户界面的示意性搜索输入屏幕的示意性描绘；

[0017] 图7根据在此示出和描述的一个或多个实施例描绘了描绘了对数据进行分析和分类的示意性方法的流程图；且

[0018] 图8根据在此示出和描述的一个或多个实施例描绘了描绘了判定私有数据是否包含非私有信息的示意性方法的流程图。

具体实施方式

[0019] 总体上参考附图，在此所述的实施例是针对用于实现隐私防火墙从而限制对位于所述隐私防火墙边界内部的服务器上所存储的私有数据语料库的访问的系统和方法。在具体实施例中，在此所述的系统和方法总体上可以被实现以保证私有数据保持安全，同时仍

然响应于一个或多个用户所提交的问题提供信息性答案，其中，所述答案是从所述私有数据获得的。一般可以假定隐私防火墙后面的所有数据都是私有数据。然而，所述数据的某些部分实际上可能包含非私有信息。另外，所述数据的包含私有信息的剩余部分还可能包含非私有元素。所述非私有信息和来自所述私有信息的所述非私有元素可以被标记为可用于在维持数据的隐私的同时回答用户所提交的问题。

[0020] 在此所披露的方法和系统可以用于例如：其中关于数据的使用是否将“跨用户”(即，实体的私有数据是否将被所述实体之外的任何人看见或推出)可能存在不确定性的实例，或其中期望提供不需要高度受控制的访问的数据储存库的情况。可以应用隐私防火墙的实例的非限制性示例包括：从人A推出的将只会影响人A并且数据储存库访问严格受控制的数据，从人A推出的可以影响人B或者数据储存库访问不受严格控制的数据，可以被映射(比如通过用户ID或互联网协议(IP)地址)至执行具体动作的用户的数据，必须遵守搜索字符串保留策略的数据(例如，必须在具体时间段内被移除或去私有化的数据)，必须遵守隐私标准的数据，以及被稍微去私有化的数据(即，用户ID或IP地址已经被移除)。

[0021] 如在此所使用的，术语“非私有信息”指包含下述信息的数据，个人或实体将将不会对所述信息有任何隐私期望。所述非私有信息可以存储在储存库中，其中，储存库中所存储的所有数据最初被假定是私有的。这样，可以确定所述数据包含非私有信息。如在此所使用的非私有信息的示意性示例是跨广泛不同的源非常常用的数据。非私有信息的另一示意性示例是不明确地与具体个人或实体相关的数据。而非私有信息的又另一示意性示例是涉及来自搜索字符串的搜索缩小构造的数据，比如人口统计、数据源、历史间隔、地理范围等。非私有信息的又另一示意性示例是涉及非私有网络浏览活动的数据，比如互联网上的任何人轻易可访问的公开信息、包含数据的公开可获得的电子文件夹等。在一些实施例中，非私有信息还可以被称为非敏感数据。在一些实施例中，对某数据是否被视为非私有的判定可以通过应用一条或多条规则完成。

[0022] 如在此所使用的，术语“私有数据”指包含下述信息的数据，个人或实体将将对所述信息有隐私期望。私有数据的示意性示例可以包括但不限于：涉及具体个人或实体的私人信息的数据，可以被映射至具体个人或实体的数据(比如包含具体用户ID、IP地址等的数据)，服从保留策略的数据，由于具体隐私标准、法规要求等而被视为私有的数据(比如被健康保险携带和责任法案(HIPAA)等视为私有的数据)，仅可以从具体个人、实体、或特定个人和/或实体分组推出的数据，可以被声称为专有的复杂数据，包含大众普遍未知信息的数据，以及将允许某人重构其中所包含的信息以得到进一步智能性(其可能危害具体个人的或实体的隐私)的数据。在一些实施例中，对某数据是否被视为私有的判定可以通过应用一条或多条规则完成。在一些实施例中，隐私防火墙后面的储存库中所存储的所有数据最初可以被假定为私有数据，直到它被分析以判定是否包括非私有信息。

[0023] 即使在此所述的储存库中的数据最初可以被视为私有的，所述数据或从所述私有数据推出的附加数据的某些部分在此可以被分类为私有数据的“非私有元素”。非私有元素的示意性示例可以包括但不限于：已经被去私有化的数据(比如已经让私人信息被从中移除的数据)，从具体数量的唯一位置出现的完全相同的数据(比如从具体数量的唯一IP地址进行的搜索)，从具体数量的唯一个人和/或实体出现的完全相同的数据，私有储存库中所存储的非私有信息，以及从私有数据获得的非标识性元数据。所述非标识性元数据可以包

括但不限于:生成了所述私有数据的个人或实体的地理区域(比如,州、省、地区、或地域,但没有比这更具体),指示生成数据的日期和小时(而不是分或秒)的时间戳,与所述私有数据相关的某些搜索词和连接器,与所述私有数据相关的市场细分,用来搜索的产品(比如具体网络浏览器、搜索引擎等),以及搜索引起的来自搜索结果的命中次数。如果被公开的数据过于具体,用于标识存留在私有储存库中的非私有信息的规则可能不完全保护用户隐私。例如,如果常见的搜索字符串被在内部公布,但它还包括完成了搜索的用户的源IP地址,用户的隐私可能被损害。

[0024] 关于用户数据,可以存在隐私频谱。例如,在一个极端,包括进行搜索的个人名字的完整搜索字符串可以被视为私有和/或敏感数据。在另一极端,仅是由于搜索时用户在某处使用了字母“e”不意味着另一个人永远不能允许字母“e”出现在所使用和公布的任何事物中。在这两个极端之前,可以存在中间立场,在所述中间立场,数据不再具有任何类型的隐私暗示。例如,如果10000个不同用户输入了搜索词“Roe v.Wade”并且然后继续浏览美国最高法院案例引用410U.S.113,如果提供了当用户开始输入“Roe v.”时直接提出跳转至410U.S.113的用户界面(UI)特征则很可能不存在对隐私的侵犯,即使要做到这一点的智能性可能与特定用户的过去动作相关。

[0025] 现在参照附图,图1根据在此所示和描绘的实施例描绘了示意性计算网络,所述示意性计算网络描绘了用于提供隐私防火墙的系统的部件,所述系统确定隐私数据储存库中的非隐私信息,基于所述非隐私信息提供对问题的响应,和/或基于所述非隐私信息自动完成搜索请求。如图1所展示的,计算机网络10可以包括广域网(WAN)(如,互联网)、局域网(LAN)、移动通信网络、公共服务电话网络(PSTN)、个人局域网(PAN)、城域网(MAN)、虚拟专用网络(VPN)和/或其他网络。计算机网络10通常可以被配置成用于电连接一个或多个计算设备和/或其部件。说明性计算设备可以包括但不局限于用户计算设备12a、服务器计算设备12b以及管理员计算设备12c。

[0026] 用户计算设备12a通常可以用作用户与连接至计算机网络10的其他部件之间的接口。因此,用户计算设备12a可以用于执行一个或多个面向用户的功能,如,接收来自用户的一个或多个输入或者将信息提供给用户,如本文中更详细描述的。另外,包括在图1中的是管理员计算设备12c。在服务器计算设备12b需要监督、更新或校正的情况下,管理员计算设备12c可以被配置成用于提供所期望的监督、更新、和/或校正。管理员计算设备12c还可以用于将附加数据输入到存储在服务器计算机设备12b上的语料库中。

[0027] 服务器计算设备12b可以从一个或多个源接收数据,存储所述数据,并将来自所述数据的某些部分的信息以对问题的答案或自动完成建议的形式提供给用户计算设备12a(当对这种信息的访问被授权并且所述信息被标记为被允许分发时)。对所述信息是否被允许分发的判定一般可以由隐私防火墙14完成,所述隐私防火墙位于所述服务器计算设备12b与所述计算机网络10之间。因而,所述隐私防火墙14(还可以被称为隐私罩)可以允许或拒绝对来自所述服务器计算设备12b处存储的数据的某信息的访问,如在此更详细描述的。

[0028] 应当理解的是,虽然用户计算设备12a和管理员计算设备12c被描绘为个人计算机并且服务器计算设备12b被描绘为服务器,但是这些是非限制性示例。更确切地,在一些实施例中,任何类型的计算设备(例如,移动计算设备、个人计算机、服务器等)可以用于这些部件中的任何部件。另外,虽然这些计算设备中的每个计算设备在图1中被展示为单件硬

件，但是这也仅是示例。更确切地，用户计算设备12a、服务器计算设备12b和管理员计算设备12c中的每一个可以表示多个计算机、服务器、数据库、部件和/或类似物。

[0029] 图2描绘了来自图1的服务器计算设备12b，进一步展示了一种用于确定非私有信息、搜索文档语料库、生成对用户所提出问题的响应、和/或生成自动完成建议的系统。另外，根据在此所示和描述的实施例，服务器计算设备12b可以包括非瞬态计算机可读介质，以用于搜索文档语料库或生成被具体化为硬件、软件、和/或固件的搜索查询。虽然在一些实施例中，服务器计算设备12b可以被配置为具有必要硬件、软件、和/或固件的通用计算机，但是在一些实施例中，所述服务器计算设备12b还可以被配置为用于执行本文所描述的功能而专门设计的专用计算机。

[0030] 还如图2所展示的，服务器计算设备12b可以包括处理器30、输入/输出硬件32、网络接口硬件34、数据存储部件36(其可以存储非私有信息38a、私有数据38b的非私有元素、以及其他数据38c)、以及非瞬态存储器部件40。存储器部件40可以被配置为易失性和/或非易失性计算机可读介质，并且如此，可以包括随机存取存储器(包括SRAM、DRAM、和/或其他类型的随机存取存储器)、闪存、寄存器、CD盘(CD)、数字通用盘(DVD)、和/或其他类型的存储部件。另外，存储器部件40可以被配置成用于存储操作逻辑42和搜索逻辑44(作为示例，所述逻辑中的每个逻辑可以被具体化为计算机程序、固件、或硬件)。本地接口46也包括在图2中并且可以被实现为总线或其他接口以便促进在服务器计算设备12b的部件之中的通信。

[0031] 处理器30可以包括被配置成用于接收和执行指令(如，来自数据存储部件36和/或存储器部件40)的任何处理部件。输入/输出硬件32可以包括监视器、键盘、鼠标、打印机、相机、麦克风、扬声器、触摸屏、和/或用于接收、发送和/或呈现数据的其他设备。网络接口硬件34可以包括任何有线或无线联网硬件，如，调制解调器、LAN端口、无线保真(Wi-Fi)卡、WiMax卡、移动通信硬件、和/或用于与其他网络和/或设备进行通信的其他硬件。

[0032] 应当理解的是，数据存储部件36可以在服务器计算设备12b本地和/或远离服务器计算设备而驻留并且可以被配置成用于存储一条或多条数据并且选择性地提供对所述一条或多条数据的访问。如图2中所展示的，数据存储部件36可以存储非私有信息38a、私有数据38b的非私有元素、以及其他数据38c，如在此更详细描述的。

[0033] 包括在存储器部件40中的是操作逻辑42和搜索逻辑44。操作逻辑42可以包括操作系统和/或用于管理服务器计算设备12b的部件的其他软件。搜索逻辑44可以被配置成用于从图形用户界面内的用户输入生成搜索查询，如下面详细描述的。

[0034] 应当理解的是，图2中所展示的部件仅是说明性的并且不旨在限制本公开的范围。更具体地，虽然图2中的部件被展示为驻留在服务器计算设备12b内，但是这是非限制性示例。在一些实施例中，所述部件中的一个或多个部件可以驻留在服务器计算设备12b外部。类似地，虽然图2涉及服务器计算设备12b，但是其他部件(诸如用户计算设备12a和管理员计算设备12c)可以包括相似的硬件、软件和/或固件。

[0035] 图3描绘了来自图1的隐私防火墙14的各个层。图3中所描绘的这些层仅是示意性的。因而，可以在不偏离本公开的范围的情况下，使用更少或额外的层。另外，某些层可以倒塌或进一步被分层为额外的层。每一层可以表示服务器计算设备12b(图1)中所包含的被提供给外部请求者(比如像用户计算设备12a的用户(图1))的数据的访问量。所述访问一般可

以是对用户所提出的问题的答案的形式或作为自动完成建议，而非对数据的直接访问。示意性层可以包括例如分布式源层20、聚合层22、隐私执行层24、以及任务特定的数据提供者层26。在一些实施例中，所述分布式源层20、所述聚合层22、所述隐私执行层24可以是高度限制性数据层，在这些层中，在不对数据进行调整的情况下此类层所分类的数据几乎不或不允许被访问，如在此更详细描述的。在一些实施例中，任务特定的数据提供者层26可以是低限制性数据层，在此层中，允许访问被所述层分类的更多或全部数据。

[0036] 分布式源层20可以对通常在例如消费者会话活动源中找到的数据进行分类。此类消费者会话活动源可以表示多个不同平台和/或应用上存在的多个源，从所述多个不同平台和/或应用接收并存储数据。作为非限制性示例，一个源可以是具体程序或应用的搜索框。来自不同源的数据可以被存储在不同数据储存库中。某些数据储存库可以比其他数据储存库具有更多数据限制。这样，不能跨所述不同储存库对所述数据进行规范化。

[0037] 聚合层22可以对通常在例如规范化的数据储存库中找到的数据进行分类。即，所述数据可以是从各数据储存库获得的，各自具有对应的本来格式，并被一个或多个规范化工具(“规范器”)规范化成单一一致的格式。在一些实施例中，当数据被规范化时，可以尽可能实用地对其进行规范化，从而使得数据中所包含的敏感信息是量被最小化。然而，在一些实施例中，所述储存库可以包含对某些模块足够的常见且一致属性集合，并且在提出具体查询(即，涉及所述数据的查询)之前完全匿名化是不可能的。因而，尽管匿名化，所述储存库中所包含的数据仍然可能是高度敏感的。相应地，对这种数据的访问会高度受限。

[0038] 隐私执行层24可以对已经经过任务特定的报告器和杀毒器模块的数据进行分类。如果以某种方式提出了查询，这种数据可以是例如包含敏感信息的经规范化数据。例如，通过以多种不同方式问多个问题并使用布尔代数运算理出比旨在揭露的更多数据，可以将隐私罩冲破。因而，可以期望严格地限制为了获得一定信息而可以提的问题类型，并且还保证被揭露的各种各样问题不能将其答案组合以泄露信息。因此，向隐私防火墙14提出的每次查询可以具有被创造用来传递结果的特定模块。每个特定模块可以在高度受限的环境中被审查和建立，从而使得所述模块充当用于将敏感数据转化成非敏感数据的桥梁。

[0039] 所述任务特定的数据提供者层26一般可以包括信息的任务特定的公共储存库，所述信息通常是非私有的被去敏感化的私有数据、或私有数据的非私有元素。这种数据可以用来回答查询。

[0040] 现在参照图4，提供了流程图，所述流程图展示了根据一个或多个实施例的实现隐私防火墙的方法。如在此更详细描述的，实施例可以使用户能够请求信息并浏览非私有信息和/或隐私信息的非私有元素。在框180，所述系统可以生成图形用户界面，用于在用户计算设备12a的显示设备上显示。所述图形用户界面被配置成使得用户可以在框182发送搜索字符串。参照图5，所述图形用户界面可以包括被配置成用于从用户接收搜索字符串的搜索字符串输入屏幕100。应该理解的是，实施例并不限于贯穿附图中所展示的图形用户界面的配置，并且其他图形用户界面配置是可能的。在一个实施例中，网络10是互联网(Internet)，并且在此所述的图形用户界面被通过浏览器呈现给所述用户。

[0041] 所述搜索字符串输入屏幕100包括搜索字符串字段103，用户可以将组成期望搜索字符串的一个或多个词语输入所述搜索字符串字段(例如，通过使用键盘)。在一个实施例中，所述搜索字符串可以是自然语言搜索字符串。例如，用户可以问比如像“最高法院关于

堕胎的标志性案例是什么?”的问题。在另一示例中,比如图5中所展示的实施例,搜索字符串“Roe v.Wade”被输入搜索字段103,因为具体用户可能对搜索与来自1973年的美国最高法院判决有关的信息。如图6中所示,在一些实施例中,用户可能仅需要在搜索字符串字段103输入一个或多个字母,并且所述系统可以基于从服务器计算设备12b内所包含的私有数据储存库获得的数据生成建议的自动完成选项。例如,如果用户正在搜索案例法数据库,当用户输入字母“RO”时,包含字母“RO”的某些自动完成选项可以被呈现给用户,比如像Roe v.Wade、In Re Ross等,如下面所示搜索字符串字段104中用户的输入。可以例如由于指示许多用户已经搜索了那些具体案件(可以可选地按照受欢迎程度排序)的数据而生成自动完成建议。因而,在用户完成输入搜索字符串之前,系统可以试图基于服务器计算设备12b中所包含的私有数据猜想用户可能想要搜索什么。然而,所获得的用于提供自动完成建议的信息一般可以是非私有信息和/或私有数据的非私有元素,从而使得可能破坏他人隐私的词语不包括在自动完成选项中。

[0042] 如图5和图6中所示,搜索字符串输入屏幕100还可以包括其他输入特征,比如选项按钮106、章节过滤输入102、和搜索发起图标105。应该理解的是,可以使用更多或更少的输入特征。在图5和图6中所展示的示例中,选项按钮106允许用户同样搜索并非用用户正搜索的语料库的本机语言的电子数据的机器翻译。还可以提供其他选项。章节过滤输入102可以使用户能够只搜索数据的具体章节或章节组合。例如,在案例法背景下,用户可以使用章节过滤输入102来只搜索案例综述章节、关键词章节、司法意见章节、事实章节等。

[0043] 用户可以通过电击或另外选择搜索发起图标105来基于被输入搜索字符串字段103的搜索字符串来发起搜索。搜索字符串中的单个词语被用作查询词语来在框184分析搜索字符串。分析搜索字符串总体上可以包括确定用户在搜索什么,这可以通过任何现在已知或以后开发的方法完成。在一些实施例中,在框186中可以基于用户的提交的搜索字符串确定合适的搜索查询。即,用户所提交的搜索字符串可以被接收并解释,并且可以基于有待访问的数据包含所述数据的系统的类型等生成合适的搜索查询。所述合适的搜索查询可以使用任何数量的查询生成技术生成。例如,可以基于对用户所提交的搜索字符串的分析生成布尔加权搜索查询。

[0044] 在框188中,可以对防火墙14后面的储存库中(比如服务器计算设备12b中)所包含的数据进行搜索。如果发现了与搜索查询有关的数据,可以对所述数据进行分析从而在步骤190中判定所述数据是私有还是非私有。在一些实施例中,可以进一步分析所述私有数据从而在步骤192中判定它是否包含非私有元素。另外,可以对所述私有数据进行分析从而在步骤194中判定某些元素是否可以被调整以使这些元素成为非私有元素。如果所述数据可以被调整,这种调整可以在步骤196中完成从而获得经调整的数据(还可以被称为已去私有化的数据)。例如,可以对所述数据进行调整从而调整时间戳和日期戳的粒度(granularity),调整地理位置的粒度等。在一些实施例中,可以在具体粒度阈值以上调整所述粒度。地理粒度阈值可以例如在州或省与城市之间(例如,阈值“以上”可以是州、省、国家、地区等粒度,并且阈值“以下”可以是附加细节比如城市、街道地址等)。时间粒度阈值可以在例如时间与分之间(例如,以小时、天、星期、月、和年显示的时间可以在阈值“以上”,而以分钟和秒显示的时间可以在阈值“以下”)。在一些实施例中,可以针对每次搜索查询完成步骤190、192、194、和196。在其他实施例中,一旦在储存库中获得了数据就可以完成步骤

190、192、194、和196，从而允许快速地完成步骤188中的搜索。

[0045] 搜索查询的查询词语用来搜索数据语料库，从而在步骤198中提供对搜索字符串的响应。所述响应总体上是从非私有信息和/或私有数据的非私有元素(包括私有数据的经调整部分(如果存在的话))推出的。所述响应可以是对用户所提问题进行响应的自然语言答案、到具体参考的一个或多个链接、供显示的已返回电子文档集合等。

[0046] 再次参照图1，在各实施例中，服务器计算设备12b可以继续收集新数据(当生成新数据和/或使得新数据可用时)。可以对所述新数据进行分析，从而使得可以关于所述数据是否包含私有或敏感数据进行判定，从而使得不允许私有或敏感信息越过防火墙14，如在此所述的。因而，图7中描绘了用于对所述数据进行分析和分类的过程。

[0047] 如图7中所示，并且还参照图1，在步骤202中可以从储存库获得所述数据。例如，在一些实施例中，所述数据可以被从远端储存库拷贝或移动至防火墙14后面的服务器计算设备12b。在其他实施例中，所述储存库可以位于服务器计算设备12b内(例如，数据存储设备36中(图2))或另外防火墙14后面，从而使得拷贝或转移是不必要的。

[0048] 如在此所述的，所述数据可以最初被假定为私有数据。在步骤204中，可以对所述数据进行分析从而在步骤206中判定所述数据是否包含非私有信息。如果所述数据不不含任何可能私有的部分，所述数据可以被标识为非私有信息并且可以在步骤208中被标识为允许在隐私防火墙14外部私用。例如，如果所述数据单单包含公开可获得的信息比如未密封的法庭记录、公开可获得的真实财产记录等，所述数据可以在步骤206中被确定为不包含私有数据并且在步骤208中被标记为允许在隐私防火墙14外部使用。

[0049] 另一方面，如果所述数据的任何一部分包含敏感、私有、或可以被看做敏感或私有的信息，所述数据可以继续被标识为私有数据。例如，如果所述数据包含法庭记录，但所述法庭记录包含隐私信息比如个人的家庭地址，所述数据可以被标识为私有信息。

[0050] 如图8中所示，对所述数据是否包含非私有信息的判定可以包括多个判定步骤。此类步骤仅仅是示意性的，并且应该理解的是，在不背离本公开范围的情况下可以完成替代、附加、或更少步骤。进一步，应该理解的是，实施例并不限于如图8中所示的这个步骤顺序。如步骤206a中所示，一个示意性步骤，可以关于所述数据是否是从位于已排除列表上的源收集的进行判定。示意性已排除列表可以包括例如包含由以下各项指定为私有的信息的列表：HIPAA、卫生保健互操作性测试和一致性协调(HITCH)工程、驾驶员隐私保护法案(DPPA)、格雷姆-里奇-比利雷法案(GLBA)(还被称为1999年金融服务现代化法案)、支付卡行业数据安全标准(PCI DSS)等。如果所述数据是从已排除列表上的源收集的，可以在步骤209中将所述数据标识为私有。如果所述数据不是从已排除列表上的源收集的，可以完成附加确定步骤，或者所述数据可以在步骤207中被标识为包含非私有信息并且在步骤208中被标记为允许在隐私防火墙外部使用。

[0051] 在示意性步骤206b中，可以关于是否在其他位置出现了完全相同的数据进行判定。例如，如果关于在搜索引擎界面输入的具体搜索字符串收集了数据，步骤206b中的确定可以看看是否从另一位置输入了相同的搜索字符串。可以通过回顾元数据等来确定位置。例如，关于具体搜索查询的元数据可以包括用来在搜索引擎界面输入搜索字符串的设备的IP地址。如果不同的IP地址输入了完全相同的搜索查询，并且此类IP地址组成不同位置(即，并非源自同一物理位置的IP地址)，可以确定所述数据已经在其他位置出现。如果所述

数据未出现在其他位置，可以在步骤209中将所述数据标识为私有。如果所述数据出现在了其他位置，可以完成附加确定步骤，或者所述数据可以在步骤207中被标识为包含非私有信息并且在步骤208中被标记为允许在隐私防火墙外部使用。

[0052] 在示意性步骤206c中，可以关于是否从至少20个不同的源获得了所述数据进行判定。例如，如果所述数据包含指示数据的源(例如，IP地址等)的元数据，所述确定可以包括回顾所述元数据以保证所述数据是从20个不同源获得的。应该理解的是，在此所使用的源数量(20)仅仅是示意性的，并且可以限定任何源数量，尤其是保证所述数据并非私有的源的数量。例如，源个数可以基于数据的类型、关于所述数据的某些规则或策略等而不同。如果所述数据未从至少20个不同源出现，可以在步骤209中将所述数据标识为私有。如果所述数据从至少20个不同源出现，可以完成附加确定步骤，或者所述数据可以在步骤207中被标识为包含非私有信息并且在步骤208中被标记为允许在隐私防火墙外部使用。

[0053] 在示意性步骤206d中，可以关于所述数据是否包含已经是公共知识的信息进行判定。例如，如果所述数据包含将一般地为私有的信息比如个人的家庭地址，但这个人已经公开地广播了他的家庭地址从而为他在他的家以外经营的业务做广告，这种信息可以被视为已经是公共知识。如果所述数据包含还不是公共知识的信息，可以在步骤209中将所述数据标识为私有。如果所述数据包含已经是公共知识的信息，可以完成附加确定步骤，或者所述数据可以在步骤207中被标识为包含非私有信息并且在步骤208中被标记为允许在隐私防火墙外部使用。

[0054] 在示意性步骤206e中，可以关于所述数据是否足够复杂以被声称为专有的进行判定。数据的复杂度可以基于数据的本质、数据源的本质、收集数据的背景、以及与数据的提供者达成的任何协议(例如，词语使用协议)逐个案例地进行。例如，如果实体开发了可以被声称为专有(比如像商业秘密)的复杂算法并且所述算法出现在数据中，至少所述数据的包含所述算法的那部分可以被视为私有的。在另一示例中，如果实体输入了被非常狭义地理解的和/或唯一的搜索字符串，所述搜索字符串可以被视为复杂的。相应地，如果所述数据被确定为足够复杂以被声称为专有的，所述数据可以在步骤209中被标识为私有数据。如果所述数据不足够复杂以被声称为专有的，可以完成附加确定步骤，或者所述数据可以在步骤207中被标识为包含非私有信息并且在步骤208中被标记为允许在隐私防火墙外部使用。

[0055] 在示意性步骤206f中，可以关于所述数据是否包含精确的时间戳进行判定。例如，如果所述数据涉及具有精确时间戳的搜索字符串(所述时间戳可以允许我们对进行了所述搜索的个人或实体进行标识)，所述数据可以是私有的。除非所述数据被适当地调整以如在此所述的调整时间戳的粒度(例如，不比搜索字符串被输入的时间更具体)，所述数据在步骤209中可以被标识为私有。如果所述数据不包含精确时间戳，或如果它已经被适当地调整，可以完成附加确定步骤，或者所述数据可以在步骤207中被标识为包含非私有信息并且在步骤208中被标记为允许在隐私防火墙外部使用。

[0056] 在示意性步骤206g中，可以关于所述数据是否包含可识别序列进行判定。可识别序列一般可以是数据串序列，所述数据串当被一起观察时包含应该私有的信息，即使孤立的所述数据串将不被视为私有。例如，如果所述数据包含可以允许我们确定与所述数据相关联的个人实体或数据的序列，这种数据可以在步骤209中被标识为私有数据。在另一示例中，如果实体使用词语“用于基于制造的集体诉讼的有利法庭”、“制作刮板小部件的公司”、

“康涅狄格(Connecticut)州内的公司”、以及“被准入康涅狄格的急诊室进行切割的个人”进行4次后续搜索,可以推出所述实体是考虑对康涅狄格的刮板小部件制造商提交集体诉讼的律师或法律事务所,这可以组成私有信息,即使所述搜索字符串单独地不会泄露这种信息并且可以不是私有。如果所述数据不包含任何可以用来标识与所述数据相关联的个人或实体的序列,可以完成附加确定步骤,或者所述数据可以在步骤207中被标识为包含非私有信息并且在步骤208中被标记为允许在隐私防火墙外部使用。

[0057] 再次参照图7,在步骤210中,可以关于所述私有数据是否包含非私有元素进行判定。即,如果所述私有数据具有某些当孤立时将组成非私有信息的部分,这种私有数据可以被标识为包含私有部分并且在步骤214中可以从中提取所述非私有元素。如本文之前所述的,可以提取非私有元素,其方式为使得不能从非私有元素搜集私有信息。例如,如果非私有元素是可以从私有数据提取的元数据,这种元数据必须在范围上被限制(例如,用户的地理位置不能比用户所在的州、地域、省等有任何更具体)。所述非私有元素在被提取时可以在步骤216中被标记为被允许在隐私防火墙14外部使用。如果所述数据不包含非私有元素,所述数据可以在步骤212中被标记为被拒绝的数据,不通过隐私防火墙14提供对所述被拒绝的数据的访问。

[0058] 应该理解的是,在此所述的实施例提供了用于提供隐私防火墙的系统和方法,所述隐私防火墙允许对私有数据的访问而不破坏(生成所述数据的)个人或实体的隐私。此处的系统和方法的本质允许用户提问题或输入搜索字符串,所述问题或搜索字符串可能需要私有数据访问以获得答案。所述用户然后可以接收对所述问题的响应或可以被提供自动完成建议,而从不获得对私有数据储存库的访问。

[0059] 虽然在此已经展示和描述了特定实施例,但应当理解的是,在不脱离所要求保护的主题的精神和范围的情况下,可作出各种其他变化和修改。此外,尽管在此已经描述了所要求保护的主题的各方面,但这些方面无需以组合使用。因此,本发明旨在所附权利要求书涵盖所有此类落入所要求保护的主题的范围内的变化和修改。

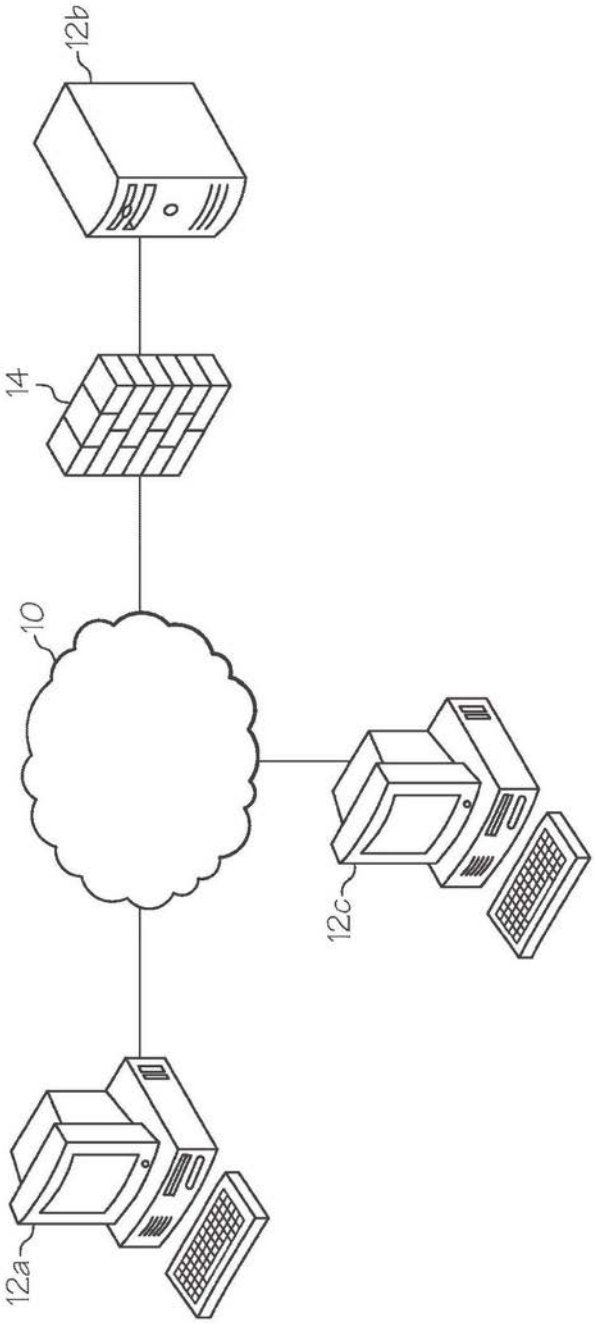


图1

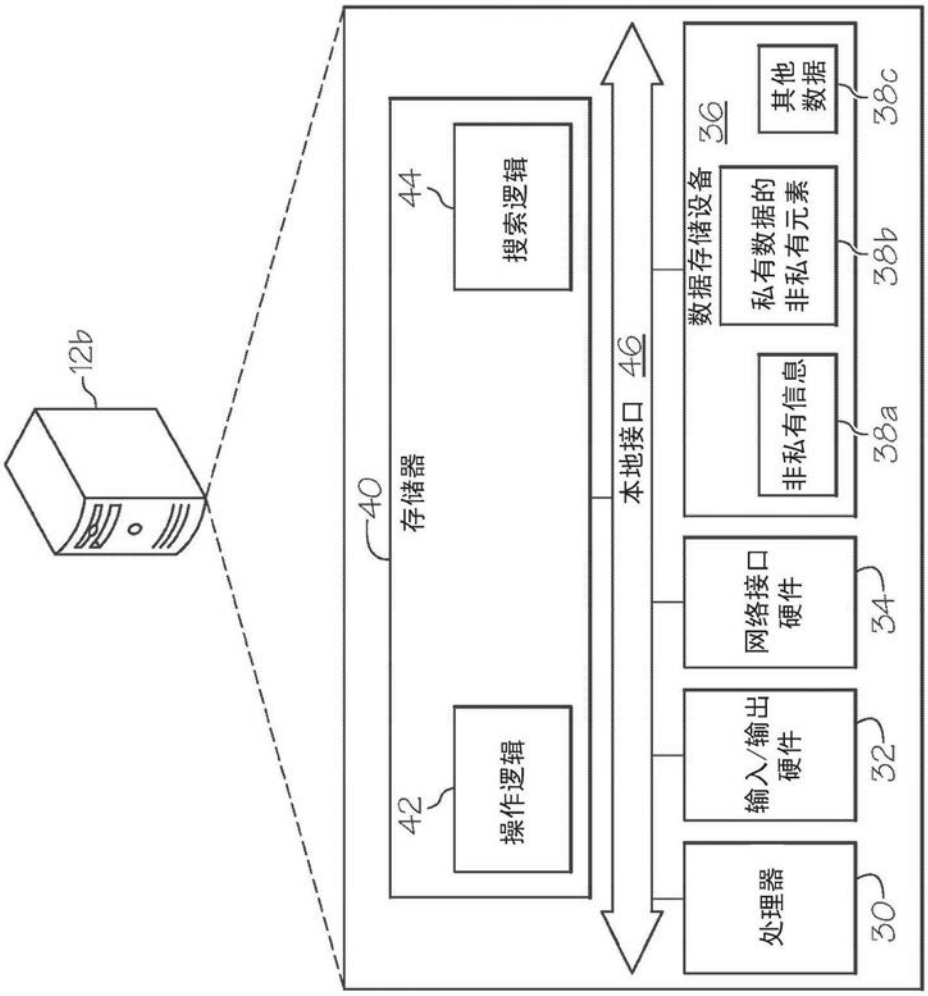

12b
40
存储器
42
操作逻辑
44
搜索逻辑
本地接口 46
处理器
30
输入/输出硬件
32
网络接口硬件
34
数据存储设备 36
非私有信息
38a
私有数据的非私有元素
38b
其他数据
38c

图2

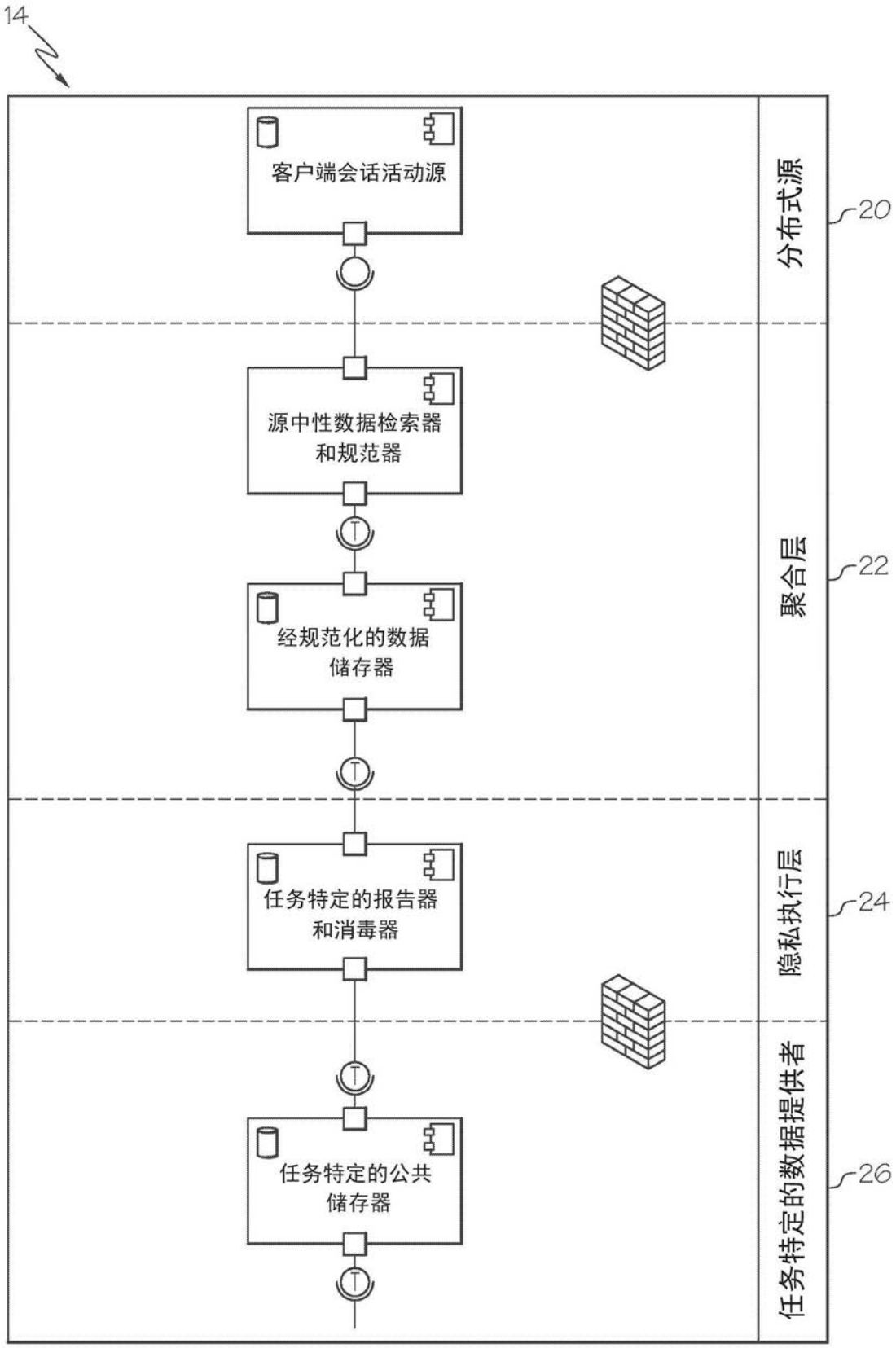

14
客户端会话活动源
源中性数据检索器
和规范器
经规范化的数据
储存器
任务特定的报告器
和消毒器
任务特定的公共
储存器
分布式源
20
聚合层
22
隐私执行层
24
任务特定的数据提供者
26

图3

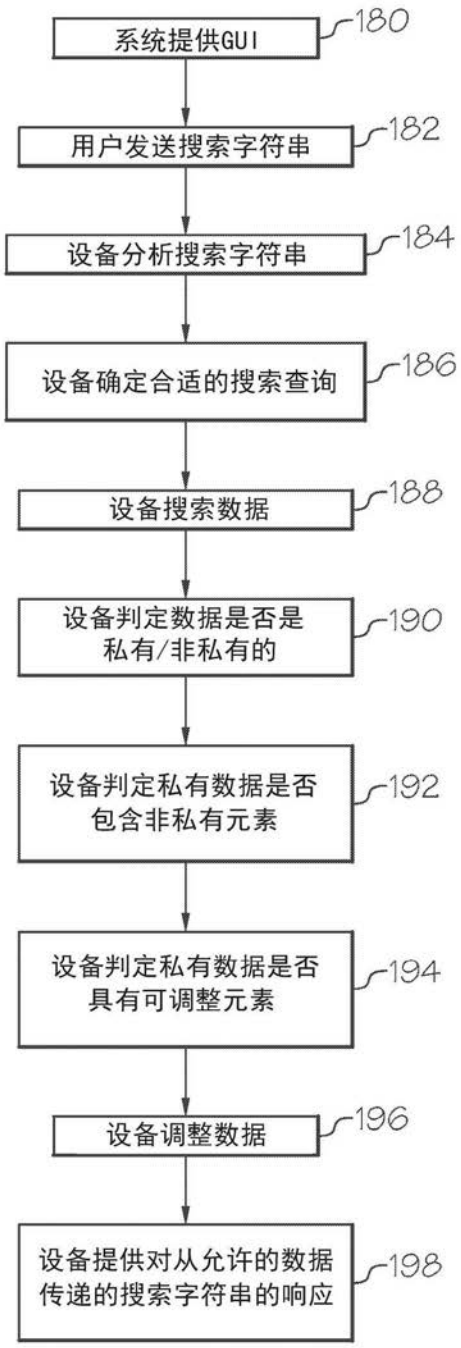

系统提供GUI
180
用户发送搜索字符串
182
设备分析搜索字符串
184
设备确定合适的搜索查询
186
设备搜索数据
188
设备判定数据是否是私有/非私有的
190
设备判定私有数据是否包含非私有元素
192
设备判定私有数据是否具有可调整元素
194
设备调整数据
196
设备提供对从允许的数据传递的搜索字符串的响应
198

图4

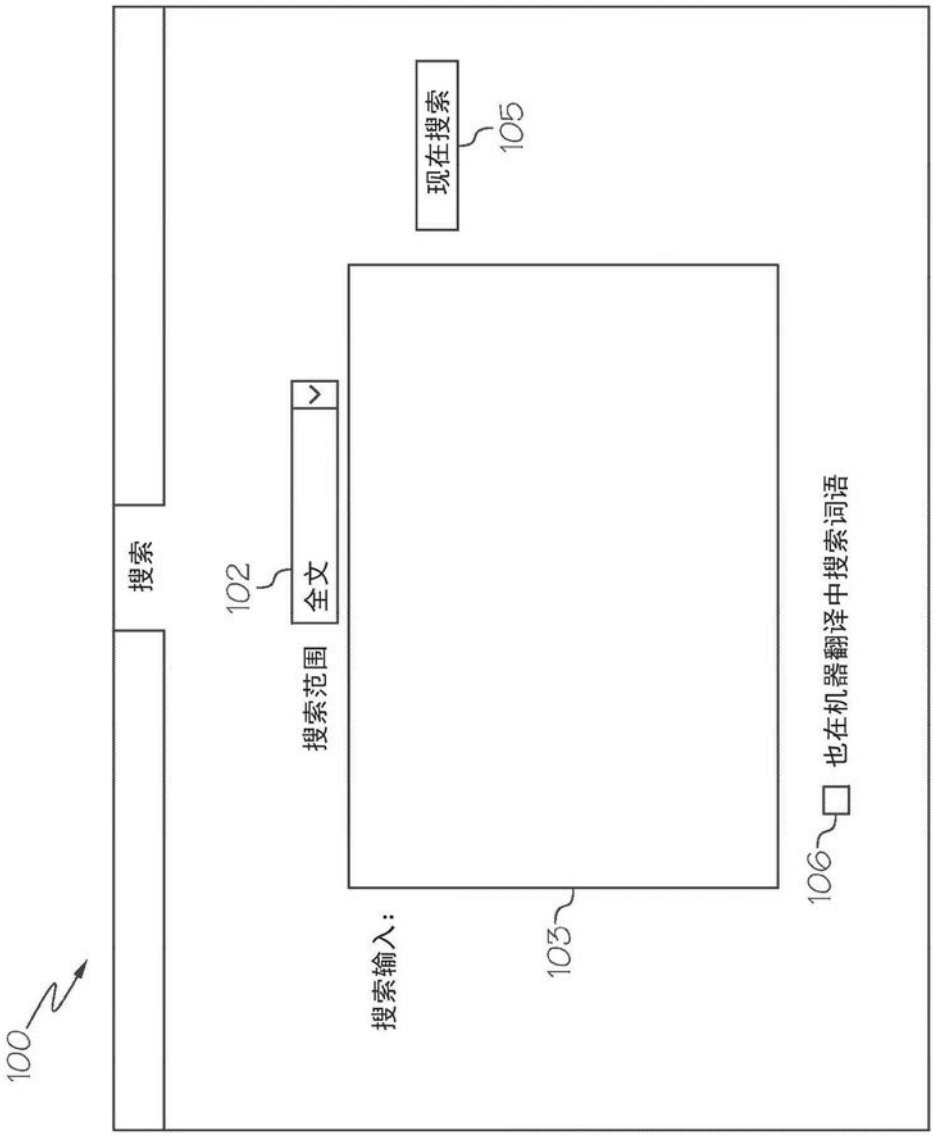


图5

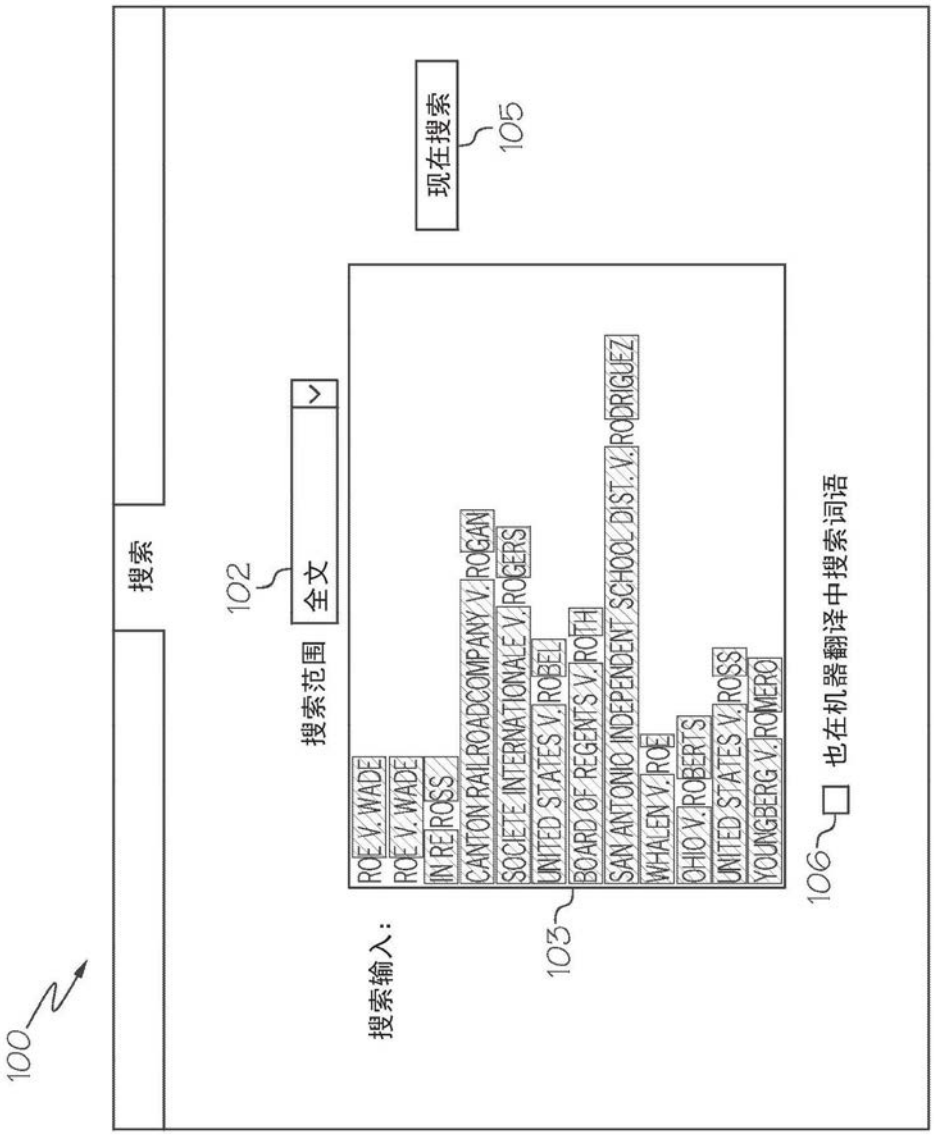

100
搜索
搜索范围
102
全文
现在搜索
105
搜索输入：
ROE V. WADE
ROE V. WADE
IN RE ROSS
CANTON RAILROAD COMPANY V. ROGAN
SOCIETE INTERNATIONALE V. ROGERS
UNITED STATES V. ROBEL
BOARD OF REGENTS V. ROTH
SAN ANTONIO INDEPENDENT SCHOOL DIST. V. RODRIGUEZ
WHALEN V. ROE
OHIO V. ROBERTS
UNITED STATES V. ROSS
YOUNGBERG V. ROMERO
103
106
也在机器翻译中搜索词语

图6

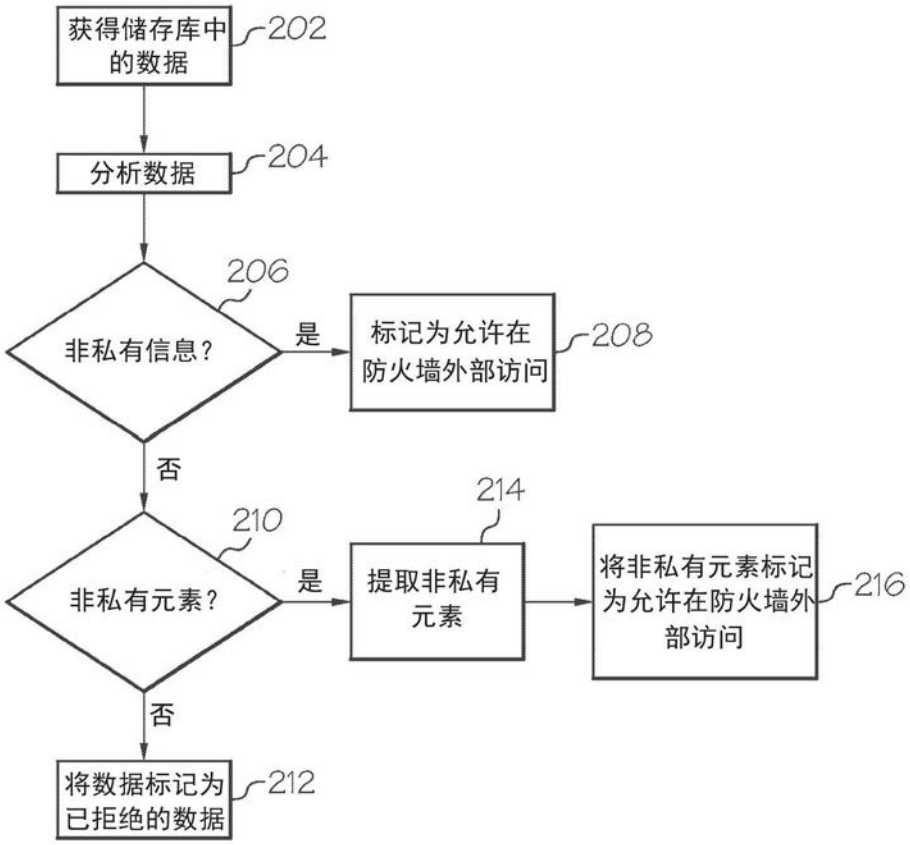

获得储存库中的数据
202
分析数据
204
206
非私有信息？
是
标记为允许在防火墙外部访问
208
否
210
非私有元素？
是
214
提取非私有元素
将非私有元素标记为允许在防火墙外部访问
216
否
将数据标记为已拒绝的数据
212

图7

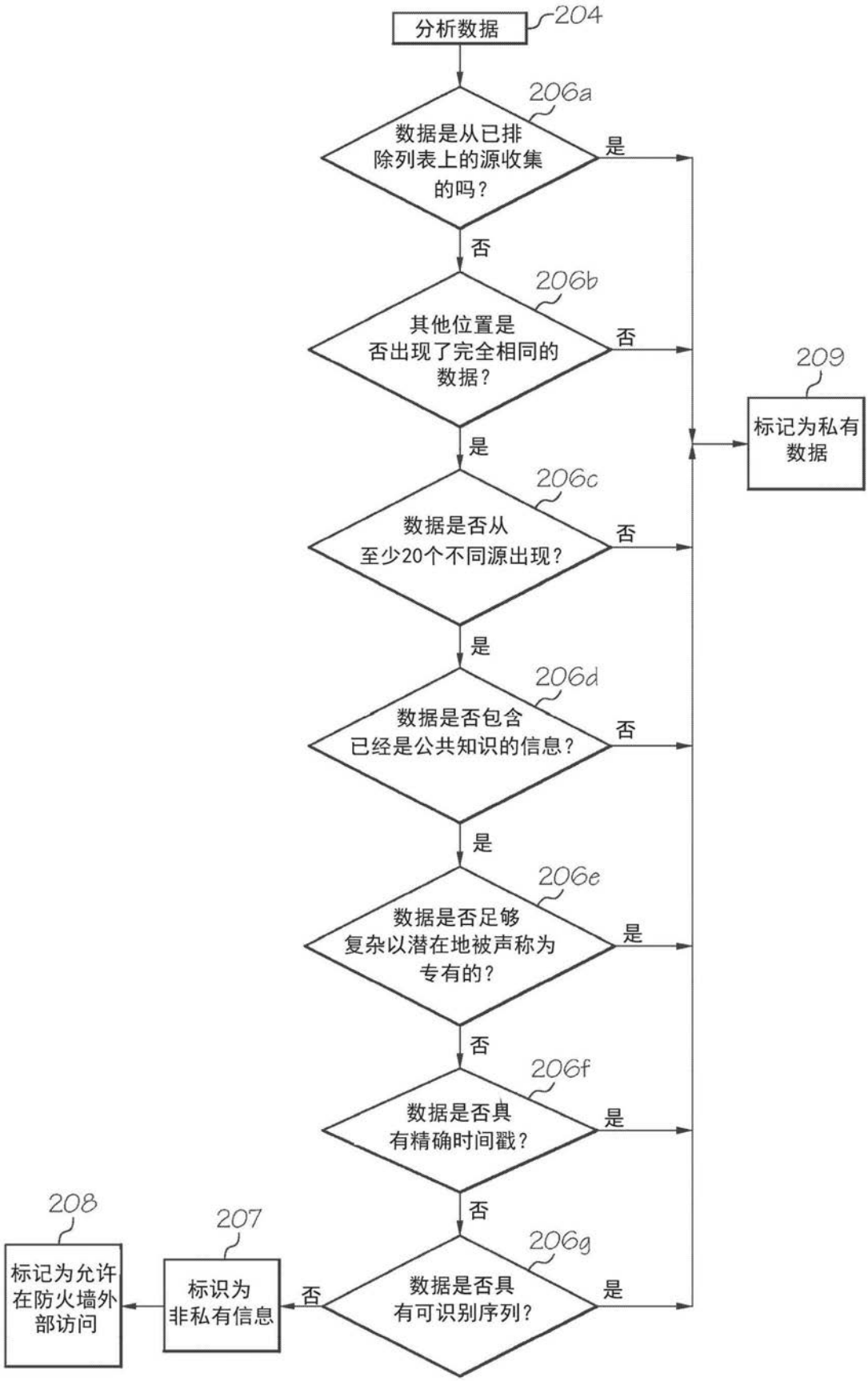

分析数据
204
206a
数据是从已排除列表上的源收集的吗？
是
否
206b
其他位置是否出现了完全相同的数据？
否
是
209
标记为私有数据
206c
数据是否从至少20个不同源出现？
否
是
206d
数据是否包含已经是公共知识的信息？
否
是
206e
数据是否足够复杂以潜在地被声称为专有的？
是
否
206f
数据是否具有精确时间戳？
是
否
206g
数据是否具有可识别序列？
是
否
207
标识为非私有信息
208
标记为允许在防火墙外部访问

图8